

GENETIC SUSCEPTIBILITY TO NDCS AND NEONATAL DNA METHYLATION

Genetic susceptibility to neurodevelopmental conditions associates with neonatal DNA methylation patterns in the general population: an individual participant data meta-analysis.

Schuurmans, I. K.^{*1,2,3}; Smajlagic, D.^{*4}; Baltramonaityte, V.⁵; Malmberg, A. L. K.⁶; Neumann, A.³; Creasey, N.^{2,3,7}; Felix, J. F.^{2,8}; Tiemeier, H.^{3,9}; Pingault, J. B.^{7,10}; Czamara, D.¹¹; Raïkkönen, K.^{6,12}; Page, C. M.¹³; Lyle R.^{13,14}; Havdahl, A.^{4,15,16}; Lahti, J.⁶; Walton, E.⁵; Bekkhus, M.⁴; Cecil, C. A. M.^{1,3,17}

*These authors have contributed equally

¹ Department of Epidemiology, Erasmus MC University Medical Center Rotterdam, Rotterdam, the Netherlands

² The Generation R Study Group, Erasmus MC University Medical Center Rotterdam, Rotterdam, the Netherlands

³ Department of Child and Adolescent Psychiatry and Psychology, Erasmus MC University Medical Center Rotterdam, Rotterdam, the Netherlands

⁴ PROMENTA Research Centre, Department of Psychology, University of Oslo, Oslo, Norway

⁵ Department of Psychology, University of Bath, Bath, United Kingdom

⁶ Department of Psychology and Logopedics, Faculty of Medicine, University of Helsinki, Helsinki, Finland

⁷ Department of Clinical, Educational, and Health Psychology, Division of Psychology & Language Sciences, University College London, London, UK.

⁸ Department of Pediatrics, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands

⁹ Department of Social and Behavioral Sciences, Harvard T.H. Chan School of Public Health, Boston, MA, USA

¹⁰ Social, Genetic & Developmental Psychiatry (SGDP) Centre, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK

¹¹ Max-Planck-Institute of Psychiatry, Department Genes and Environment, Munich, Germany.

¹² Department of Obstetrics and Gynecology, Helsinki University Hospital and University of Helsinki, Helsinki, Finland

¹³ Centre for Fertility and Health, Norwegian Institute of Public Health, Oslo, Norway

¹⁴ Department of Medical Genetics, Oslo University Hospital, Oslo, Norway

¹⁵ PsychGen Centre for Genetic Epidemiology and Mental Health, Norwegian Institute of Public Health, Oslo, Norway

¹⁶ Nic Waals Institute, Lovisenberg Diaconal Hospital, Oslo, Norway

¹⁷ Molecular Epidemiology, Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, the Netherlands

Words abstract: 300

Words main text: 3,497

References: 30

Tables and figures: 3 tables, 3 figures

Short title: Genetic susceptibility to NDCs and neonatal DNA methylation

Correspondence: Charlotte A.M. Cecil, Department of Child and Adolescent Psychiatry and Psychology, Erasmus MC University Medical Center Rotterdam, PO Box 2040, 3000 CA Rotterdam, the Netherlands. Email: c.cecil@erasmusmc.nl

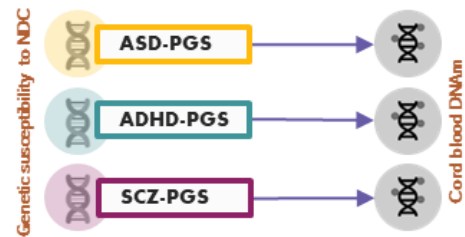
NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

Genetic susceptibility to neurodevelopmental conditions (NDCs) associates with neonatal DNA methylation patterns in the general population: *an individual participant data meta-analysis.*

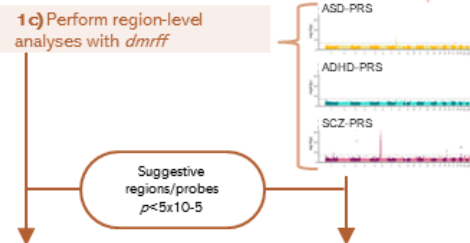
1 EWAS of genetic susceptibility for NDCs and DNA methylation at birth



1a) 3 EWASs run per subcohort with ASD-PGS, ADHD-PGS, and SCZ-PGS as predictors using *robust linear regression*



1b) Meta-analyze subcohort results with *METAL* (probe-level)

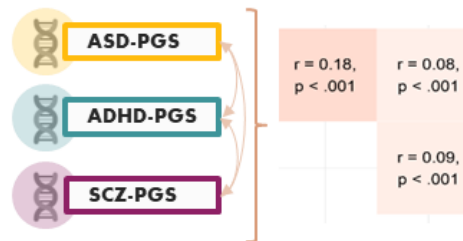


1d) Functional characterization of suggestive probes/regions

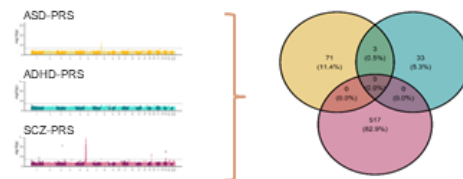


2 Cross-NDC comparisons

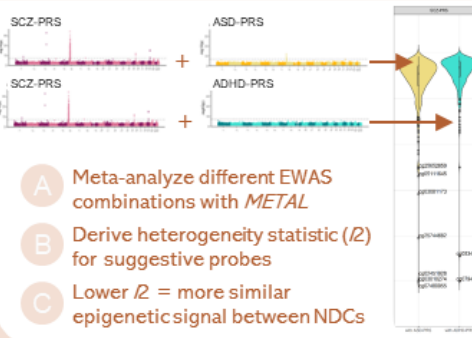
2a) Correlation between ASD-PGS, ADHD-PGS, and SCZ-PGS



2b) Identifying suggestive probes ($p < 5 \times 10^{-5}$) associated with 2 or more NDCs

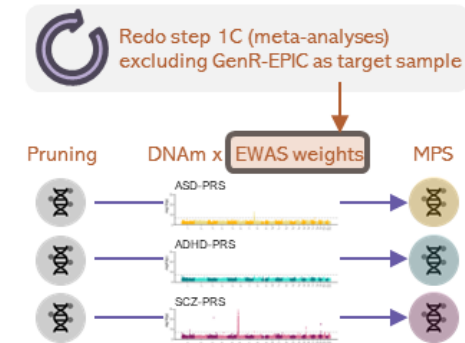


2c) Examining heterogeneity in PRS-associated DNAm patterns, by meta-analyzing different NDC-pairs

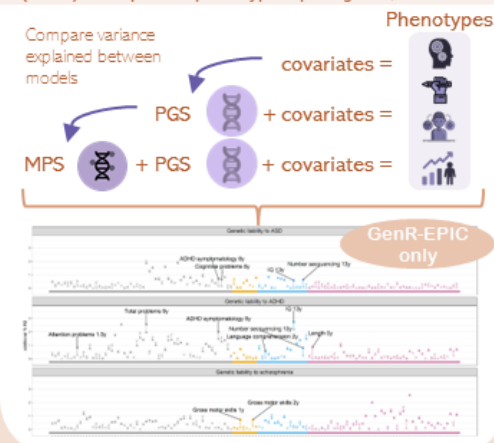


3 Phenotypic associations

3a) Build a DNAm-based measure of genetic susceptibility, using a methylation profile score (MPS) derived from EWAS



3b) Examine incremental variance explained by MPS in (neuro)developmental phenotypes up to age 14, over PGS



GENETIC SUSCEPTIBILITY TO NDCS AND NEONATAL DNA METHYLATION

Abstract

Background. Autism spectrum disorder (ASD), attention-deficit/hyperactivity disorder (ADHD), and schizophrenia (SCZ) are highly heritable and linked to disruptions in foetal (neuro)development. While epigenetic processes are considered an important underlying pathway between genetic susceptibility and neurodevelopmental conditions, it is unclear (i) whether genetic susceptibility to these conditions is associated with epigenetic patterns, specifically DNA methylation (DNAm), already at birth; (ii) to what extent DNAm patterns are unique or shared across conditions, and (iii) whether these neonatal DNAm patterns can be leveraged to enhance genetic prediction of (neuro)developmental outcomes.

Methods. We conducted epigenome-wide meta-analyses of genetic susceptibility to ASD, ADHD, and schizophrenia, quantified using polygenic scores (PGSs) on cord blood DNAm, using four population-based cohorts ($n_{\text{pooled}}=5,802$), all North European. Heterogeneity statistics were used to estimate overlap in DNAm patterns between PGSs. Subsequently, DNAm-based measures of PGSs were built in a target sample, and used as predictors to test incremental variance explained over PGS in 130 (neuro)developmental outcomes spanning birth to 14 years.

Outcomes. In probe-level analyses, SCZ-PGS associated with neonatal DNAm at 246 loci ($p < 9 \times 10^{-8}$), predominantly in the major histocompatibility complex. Functional characterization of these DNAm loci confirmed strong genetic effects, significant blood-brain concordance and enrichment for immune-related pathways. 8 loci were identified for ASD-PGS (mapping to *FDFT1* and *MFHAS1*), and none for ADHD-PGS. Regional analyses indicated a large number of differentially methylated regions for all PGSs (SCZ-PGS: 157, ASD-PGS: 130, ADHD-PGS: 166). DNAm signals showed little overlap between PGSs. We found suggestive evidence that incorporating DNAm-based measures of genetic susceptibility at birth increases explained variance for several child cognitive and motor outcomes over and above PGS.

Interpretation. Genetic susceptibility for neurodevelopmental conditions, particularly schizophrenia, is detectable in cord blood DNAm at birth in a population-based sample, with largely distinct DNAm patterns between PGSs. These findings support an early-origins perspective on schizophrenia.

Funding. HorizonEurope; European Research Council

Keywords. Population-based; Genetic susceptibility; DNA methylation; Epigenetics; Neurodevelopmental conditions; Generation R Study; PREDO; ALSPAC; MoBa

GENETIC SUSCEPTIBILITY TO NDCS AND NEONATAL DNA METHYLATION

Introduction

Neurodevelopmental conditions (NDCs) refer to a set of complex, multifactorial conditions involving perturbations in brain development that begin in foetal life.¹ The corresponding ICD-11 diagnostic category includes in this category conditions with a developmental onset, such as autism spectrum disorder (ASD) and attention-deficit/ hyperactivity disorder (ADHD). Schizophrenia is also regarded as having neurodevelopmental origins, despite its later age of onset.^{1,2} A common feature of these three conditions is their high genetic contribution, evidenced by both family-based studies (twin-based heritability estimates ~80%)³⁻⁵ as well as by large-scale genome-wide association studies (GWAS; single nucleotide polymorphism-based heritability ~20-40%).⁶ However, the biological mechanisms underlying phenotypic presentation of these disorders remain poorly understood. Epigenetic processes that modulate gene expression, such as DNA methylation (DNAm), may be promising molecular candidates in search for biological markers and mediators of genetic and environmental influences on neurodevelopmental risk.

DNAm from peripheral blood is increasingly being used in the clinic together with genetic data for diagnosing Mendelian neurodevelopmental disorders, showing utility in differentiating complex cases with ambiguous phenotypes, compared to the use of genetic data alone.⁷ In contrast, the extent to which genetic susceptibility to *complex* (highly polygenic) NDCs associates with DNAm is less clear. So far, a limited set of studies have examined whether GWAS-derived polygenic scores (PGS) of ASD, ADHD, and schizophrenia associate with DNAm patterns, irrespective of diagnostic status (all case-control, $n < 1,300$).⁸⁻¹⁰ Using a PGS of ASD, Hannon et al. (2018) identified genome-wide significant associations with DNAm from neonatal heel-pricks at two CpGs⁸, Mooney et al (2020) found that a PGS of ADHD was associated with saliva-derived DNAm in mid-to-late childhood at one CpG. Finally, a PGS of schizophrenia was found to associate with DNAm from adult blood at two CpGs, but those findings were not replicated in an independent cohort.¹⁰ For ASD and ADHD, DNAm patterns more strongly associated with PGS than with diagnostic status itself.^{8,9}

While these studies provide preliminary support for a link between genetic susceptibility to NDCs and DNAm, key gaps remain. First, studies have focused exclusively on clinical case-control samples⁸⁻¹⁰ and associations within the *general population* have yet to be characterized. This is important given that NDCs typically manifest as a continuum, with a diagnosis of ADHD, ASD, or SCZ representing the tail end of this continuum. Furthermore, DNAm patterns are developmentally dynamic and tissue-specific,¹¹ and existing studies have varied in the age and tissue of DNAm assessment (i.e., neonatal heel-pricks, saliva in childhood, blood in adulthood). Growing evidence suggests that DNAm variation at birth may be a particularly informative marker of neurodevelopmental risk, with several recent studies identifying cord blood DNAm as a stronger predictor of neurodevelopmental risk than DNAm measured in childhood. Cord blood DNAm may thus be a better proxy for congenital effects associated with NDCs, with this signal becoming noisier over time (e.g., due to postnatal exposures and immune-related changes¹¹). In addition, the fact that cord blood DNAm precedes symptom onset makes it an especially promising tissue for early risk prediction, while also minimizing reverse causality. Second, studies have so far focused on PGS for individual NDCs in isolation, despite evidence of genetic and phenotypic overlap between these susceptibilities.⁶ Examining multiple PGS within the same set of individuals would offer a valuable opportunity to characterize unique versus shared epigenetic correlates of genetic risk across conditions. Finally, no research has examined the potential utility of the identified epigenetic marks in predicting (neuro)developmental outcomes. Given that PGS currently explain little variance in NDCs in the general paediatric population, examining whether incorporating additional information on genetic susceptibility

GENETIC SUSCEPTIBILITY TO NDCS AND NEONATAL DNA METHYLATION

from a different biological and regulatory level amplifies PGS prediction could have important implications for early risk detection.¹²

To address these gaps, we conducted a large-scale epigenome-wide association meta-analysis of genetic susceptibility to ASD (ASD-PGS), ADHD (ADHD-PGS), and schizophrenia (SCZ-PGS), leveraging individual participant data from four population-based prospective birth cohorts with DNAm obtained in the same tissue and time point (cord blood at birth) with a total combined sample size of 5,802 participants from North European datasets. Specifically, (i) we investigated epigenome-wide associations of PGSs with cord blood DNAm in the general population (probe and region level), and performed follow-up characterization to examine genetic influences (i.e., methylation quantitative trait loci (mQTL) and twin heritability estimates), associations with gene expression, blood-brain concordance, enriched functional pathways, and developmental dynamics of identified signals; (ii) we examined whether epigenetic patterns are distinct or shared across genetic susceptibilities, and (iii) we explored whether incorporating a DNAm-based measure of genetic susceptibility at birth amplifies PGS prediction of neurodevelopmental outcomes across childhood.

Methods

Study population

This study features four population-based birth cohorts: the Generation R Study (GenR),¹³ The Prediction and prevention of preeclampsia and intrauterine growth restriction (PREDO),¹⁴ the Avon Longitudinal Study of Parents and Children (ALSPAC),¹⁵ and the Norwegian Mother, Father, and Child Cohort Study (MoBa).¹⁶ Inclusion criteria and cohort-specific descriptions of methods can be found in the **Supplementary Information**. The meta-analyses included 5,802 participants of North European ancestry (**Table 1**). This study was conducted according to the Helsinki Declaration of the World Medical Association, and written informed consent was provided by all participating mothers.

Genetic susceptibility for NDCs

We calculated polygenic scores with the latest GWAS summary statistics for three NDCs: ASD (ASD-PGS), ADHD (ADHD-PGS), and schizophrenia (SCZ-PGS)¹⁷⁻¹⁹, using *PRSice2* (<https://choishingwan.github.io/PRSice/>). First, we clumped correlated SNPs within a 250 kb window at an R^2 threshold of 0.1. Second, PGSs were thresholded, by calculating PGSs against multiple p -value thresholds (only SNPs with a GWAS p -value below threshold are included in the PGS), and selecting the threshold for which PGS explains most variance in diagnosis-related measures across cohorts (0.5 for ASD-PGS and 0.01 for ADHD-PGS). For SCZ-PGS, a fixed threshold of 0.05 was chosen in accordance with the original GWAS,¹⁷ due to the lack of relevant phenotypic measures in several of the cohorts. Detailed descriptions of genotyping and PGS calculation are available in the **Supplementary Information**.

DNA methylation

DNAm was extracted from cord blood and bisulfite-converted with the EZ-96 DNA Methylation kit (Zymo Research Corporation, Irvine, USA). Samples were run on the Illumina Infinium HumanMethylation450 BeadChip (Illumina Inc., San Diego, USA) or the Illumina MethylationEPIC BeadChip (EPIC), which include 485,577 and 867,531 CpGs, respectively. DNAm beta values were winsorised ($> \pm 3$ SD) to reduce the influence of outliers. We excluded sites only available in one cohort and sites that are cross-reactive or polymorphic (indicated by the R package *maxprobes*; <https://github.com/markgene/maxprobes>), leaving

Table 1. Population characteristics for all sub-cohorts cohorts separately.

	GenR _{450K}	GenR _{EPIC}	PREDO	ALSPAC	MoBa-1	MoBa-2	MoBa-4	MoBa-8
Cohort characteristics								
Cohort	GenR	GenR	PREDO	ALSPAC	MoBa	MoBa	MoBa	MoBa
Country	NL	NL	FI	UK	NO	NO	NO	NO
Sample size	1,317	1,097	767	731	355	128	739	668
DNAm array for cord blood	450K	EPIC	450K	450K	450K	450K	EPIC	EPIC
Child characteristics								
Child sex, female, n (%)	657 (49·1)	568 (51·8)	362 (47·2)	372 (50·9)	169 (47·5)	70 (53·9)	362 (51·3)	353 (54·6)
Gestational age at birth, weeks, mean $\pm$ SD	40·1 $\pm$ 1·5	40·0 $\pm$ 1·5	39·8 $\pm$ 1·6	39·6 $\pm$ 1·5	39·6 $\pm$ 1·5	39·4 $\pm$ 1·5	39·6 $\pm$ 1·7	39·6 $\pm$ 1·9
Maternal characteristics								
Age, years, mean $\pm$ SD	31·7 $\pm$ 4·2	31·4 $\pm$ 4·3	33·3 $\pm$ 5·8	29·8 $\pm$ 4·5				
Self-reported smoking during pregnancy, yes, n (%)	275 (23·0)	227 (22·9)	31 (4·0)	75 (10·3)	40 (11·1)	12 (10·9)	55 (9·6)	67 (12·2)
Educational level, low, n (%)	24 (1·8)	45 (4·2)	335 (43·7)	109 (15·1)	6 (1·8)	4 (3·3)	12 (1·8)	8 (1·3)
Educational level, medium, n (%)	426 (32·8)	377 (35·3)	173 (22·6)	457 (63·3)	133 (33·3)	33 (27·3)	225 (33·4)	199 (32·3)
Educational level, high, n (%)	848 (65·3)	645 (60·4)	239 (31·2)	156 (21·6)	220 (64·9)	84 (69·4)	437 (64·8)	410 (66·5)

Note. NL = The Netherlands; UK = United Kingdom; FI = Finland; NO = Norway; 450K = Illumina Infinium HumanMethylation450 BeadChip; EPIC = Illumina MethylationEPIC BeadChip. There were missing values maternal characteristics variables. Self-reported smoking is shown here as more directly interpretable than prenatal maternal smoking as assessed by DNAm. Rates of low/medium/high education are not directly comparable across cohorts as educational systems differ between countries. See Supplementary Information for specific definitions.

GENETIC SUSCEPTIBILITY TO NDCS AND NEONATAL DNA METHYLATION

795,580 sites (380,778 EPIC-only, $n=2,504$). Cohort-specific quality control and normalization procedures are described in the **Supplementary Information**.

Covariates

Covariates included sex, gestational age at birth, prenatal maternal smoking assessed by DNAm (for comparability across cohorts), cell-type proportions estimated via the combined cord-blood reference panel,²⁰ genomic principal components to adjust for population stratification, and technical covariates to adjust for batch effects (differing per cohort; for full details, see the **Supplementary Information**).

Analyses

Step 1. Epigenome-wide associations. Each cohort performed a probe-level epigenome-wide association study (EWAS) to assess associations between PGSs (ASD-PGS, ADHD-PGS, SCZ-PGS; predictors) and DNAm at birth (outcome) with covariate adjustment, separately for each NDC and CpG. We run robust linear regressions (iterated re-weighted least squares), which is less sensitive to potential heteroscedasticity and influential outliers, using the *MASS* package (version 7.3) in R version 4.1.3. Findings from individual cohorts were pooled with inverse-variance weighted fixed effects meta-analysis with *METAL* (EWAS-MA; <http://www.sph.umich.edu/csg/abecasis/metal/>). To assess the stability of probe-level results, a leave-one-out meta-analysis on sub-cohort was performed for the top 10 significant hits per PGS. In addition, we performed regional analyses, examining differentially methylated regions (DMRs) with the *dmrff* R package, which defines significant DMRs based on proximity, p-value, and directional consistency of CpGs (default settings). Associations were defined as genome-wide significant at a threshold of $p < 9 \times 10^{-8}$ (Bonferroni-corrected for number of effective tests)²¹ and suggestive at $p < 5 \times 10^{-5}$. Suggestive results were annotated using the *meffil* R package, and functionally characterized using several openly available resources (**Table 2**).

Step 2. Cross-NDC comparisons. We first examined correlations between ASD-PGS, ADHD-PGS, and SCZ-PGS. Second, we identified CpGs shared across the PGS-specific EWAS-MA results, defined as CpGs showing suggestive associations with >2 PGSs. Third, we pooled together results from each EWAS-MA in pairs (e.g. EWAS-MA ASD-PGS with EWAS-MA ADHD-PGS) with inverse-variance weighted fixed effects meta-analysis (cross-NDC meta-analyses) in *METAL*. We examined heterogeneity (I^2) statistics for suggestive sites, which quantifies the proportion of variance across PGS-specific EWAS-MA results attributable to heterogeneity rather than chance. Lower I^2 values reflect epigenetic associations across the compared PGSs are more similar.

Step 3. Phenotypic associations. Finally, we built a DNAm-based measure of genetic susceptibility, by constructing methylation profile scores (MPS) for each PGS at birth (capturing the broader epigenetic signal associated with genetic susceptibility for ASD, ADHD, and schizophrenia), and tested whether these MPSs explain additional variance in (neuro)developmental outcomes over and above the PGSs themselves. To avoid over-fitting, we re-ran the EWAS-MAs removing one dataset, which we used as a target sample (GenRE_{EPIC}, $n=1,097$). We multiplied the EWAS-MA weights for suggestive sites ($p < 5 \times 10^{-5}$) with methylation beta-values in GenRE_{EPIC}, performed clumping based on co-methylation patterns within GenRE_{EPIC}, and aggregated weighted sites into a single aggregate score similar to PGS calculation. As a baseline, we examined the incremental variance explained by PGS, above covariates, in (neuro)developmental phenotypes (spanning motor, cognitive, and behavioural domains at different time points; anthropometric variables included for comparison). Next, we evaluated incremental variance explained by MPS, above PGS and covariates. Significance was defined

Table 2. Functional characterization of probe-level suggestive sites ($p < 5 \times 10^{-5}$).

	ASD-PRS (71 suggestive sites)	ADHD-PRS (36 suggestive sites)	SCZ-PRS (517 suggestive sites)	SCZ-PRS –MHC (196 suggestive hits)
Methylation quantitative trait loci (mQTL) through the GoDMC database (http://mqtl.db.godmc.org.uk/)	24 sites (32%), significantly more than background ($p=0.001$)	12 sites (33%), not significant ($p=0.270$)	379 sites (73%), significantly more than background ($p=3.0 \times 10^{-69}$)	128 sites (64%), significantly more than background ($p=5.9 \times 10^{-24}$)
Average twin-heritability estimates (http://www.epigenomicslab.com/online-data-resources)	Average twin heritability of 30% [range 0-86%], significantly more than background ($p=0.016$)	Average twin heritability of 38% [range 0-89%], significantly more than background ($p=0.006$)	Average twin heritability of 51% [range 0-99%], significantly more than background ($p=1.3 \times 10^{-72}$)	Average twin heritability of 50% [range 0-99%], significantly more than background ($p=1.5 \times 10^{-28}$)
HELIX Web Catalogue to test whether the identified top hits are associated with gene expression changes in blood by expression quantitative trait methylation (eQTM mapping) (https://helixomics.isglobal.org/)	7 sites (9%), significantly more than background ($p=0.001$)	4 sites (11%), significantly more than background ($p=0.016$)	203 sites (39%), significantly more than background ($p=6.2 \times 10^{-140}$)	56 sites (28%), significantly more than background ($p=1.7 \times 10^{-34}$)
Cross-tissue correspondence tools to probe blood-brain concordance of the identified sites (BECon) (https://redgar598.shinyapps.io/BECon/)	average blood-brain correlation of 0.04 [range -0.45-0.56]	average blood-brain correlation of 0.01 [range -0.52-0.30]	average blood-brain correlation of 0.11 [range -0.56-0.85]	average blood-brain correlation of 0.07 [range -0.41-0.8]
MissMethyl package (version 1.34.0) to identify enrichment for broader molecular pathways and functions (GO collection) (for full list, see Supplementary Table 12-13)	No enrichment	No enrichment	166 pathways, mainly adaptive immune system; 45 when restricted to epigenome-wide significant hits ($p < 9 \times 10^{-8}$)	No enrichment
EpiDelta tool, which characterizes longitudinal epigenetic changes over the first two decades of life (http://epidelta.mrcieu.ac.uk/), checked for i) increase or decrease in methylation across childhood, ii) non-linearity of change, or iii) differences in interindividual variability	No significant differences	No significant differences	As compared to background, significantly more non-linear changes in DNAm across childhood ($p=1.0 \times 10^{-34}$)	As compared to background, significantly more non-linear changes in DNAm across childhood ($p=4.0 \times 10^{-6}$)
EWAS Catalogue to see if hits were identified in association to other phenotypes in previous EWASs (https://www.ewascatalog.org/). Here we only list notable traits, for full list see Supplementary Table 14).	Health conditions, protein levels, ADHD, smoking, alcohol consumption	Health conditions, protein levels	Health conditions, protein levels, several psychiatric conditions (e.g., schizophrenia, ADHD, and depression), neurocognitive conditions (e.g., as mild cognitive impairment and Alzheimer's disease), and adversities (e.g., exposure to child abuse).	Health conditions, protein levels, several psychiatric conditions (e.g., schizophrenia, ADHD, and depression), neurocognitive conditions (e.g., as mild cognitive impairment and Alzheimer's disease), and adversities (e.g., exposure to child abuse).

Note. Functional enrichment for suggestive probes was compared to background CpGs (array-wide DNAm) using a Fisher's exact test. We note that all of these tools, except missMethyl, are only available for probes on the 450K array. More extensive information on functional characterization is available in Supplementary Table 1-3.

GENETIC SUSCEPTIBILITY TO NDCS AND NEONATAL DNA METHYLATION

phenotype-wide as $p < 8 \times 10^{-4}$ (Bonferroni-corrected for the number of effective tests $n=61$, Galwey method) and nominally at $p < 0.05$.

Further details about step 1-3 are described in the **Supplementary Information**.

Results

Study characteristics

A total of 5,802 participants (50.2% female) were included in this study (**Table 1**). Gestational age at birth was similar across cohorts, however, rates of self-reported pregnancy smoking (any) ranged from 9.6% in MoBa-4 to 23.0% in GenR.

Epigenome-wide associations of genetic susceptibility for NDCs and DNA methylation at birth

In probe-level analyses, ASD-PGS was associated with DNAm at 8 CpGs at birth after Bonferroni-correction ($p < 9 \times 10^{-8}$; in total 74 suggestive at $p < 5 \times 10^{-5}$), all of which were located on chromosome 8 close to the *MFHAS1* and *FDT1* genes (**Supplementary Table 1**), and nearly all present only on the EPIC array. No probe-level hits were identified for ADHD-PGS after Bonferroni-correction (36 suggestive, **Supplementary Table 2**). In contrast, the SCZ-PGS was associated with DNAm at 246 CpGs after Bonferroni-correction (517 suggestive, **Supplementary Table 3**). The majority of these sites were on chromosome 6 (96% at $p < 9 \times 10^{-8}$, 87% at $p < 5 \times 10^{-5}$), mostly within the major histocompatibility complex (MHC; between positions 29,640,000 and 33,120,000; 62% of Bonferroni-correction hits, 61% suggestive hits). Consistently, probe-level associations with SCZ-PGS were attenuated when the MHC was partially or fully omitted from the PGS (sensitivity analyses in GenR_{EPIC}, see **Supplementary Information**). **Table 3** shows the 10 top-ranking CpGs for each PGS. EWAS-MA results showed no indication of genomic inflation (**Figure 1**), and leave-one-out results indicated that associations were unlikely driven by a single cohort (**Supplementary Figure 1**).

In region-level analyses (combining proximal CpGs into a smaller set of DNAm regions), a large number of DMRs were identified for all three PGSs, with 130 regions associated with ASD-PGS ($p < 9 \times 10^{-8}$; 251 at $p < 5 \times 10^{-5}$), 166 regions with ADHD-PGS (305 suggestive), and 157 regions with SCZ-PGS (297 suggestive, **Supplementary Table 4-6**). Regions overlapped only slightly with probe-level results (genes overlapping between probe-level and region-level results: ASD-PGS 5.4%; ADHD-PGS 7.4%; SCZ-PGS 12.4%).

Follow-up on probe-level and region-level analyses indicated that SCZ-PGS sites, as compared to ASD-PGS and ADHD-PGS sites, showed (i) more genetic influence (more mQTLs, higher twin heritability estimates), (ii) stronger blood-brain concordance, (iii) more enriched biological pathways (particularly related to adaptive immune response), (iv) more significant non-linear change across childhood, and (v) more often reported links to psychiatric disorders, neurocognitive conditions, and adversities, based on existing EWAS studies. These patterns were largely driven by the MHC locus because results attenuated when analyses were restricted to suggestive SCZ-PGS hits outside this locus. Full functional characterization of probe-level results can be found in **Table 2** and **Supplementary Table 1-3**; for region-level functional characterization see **Supplementary Table 7**.

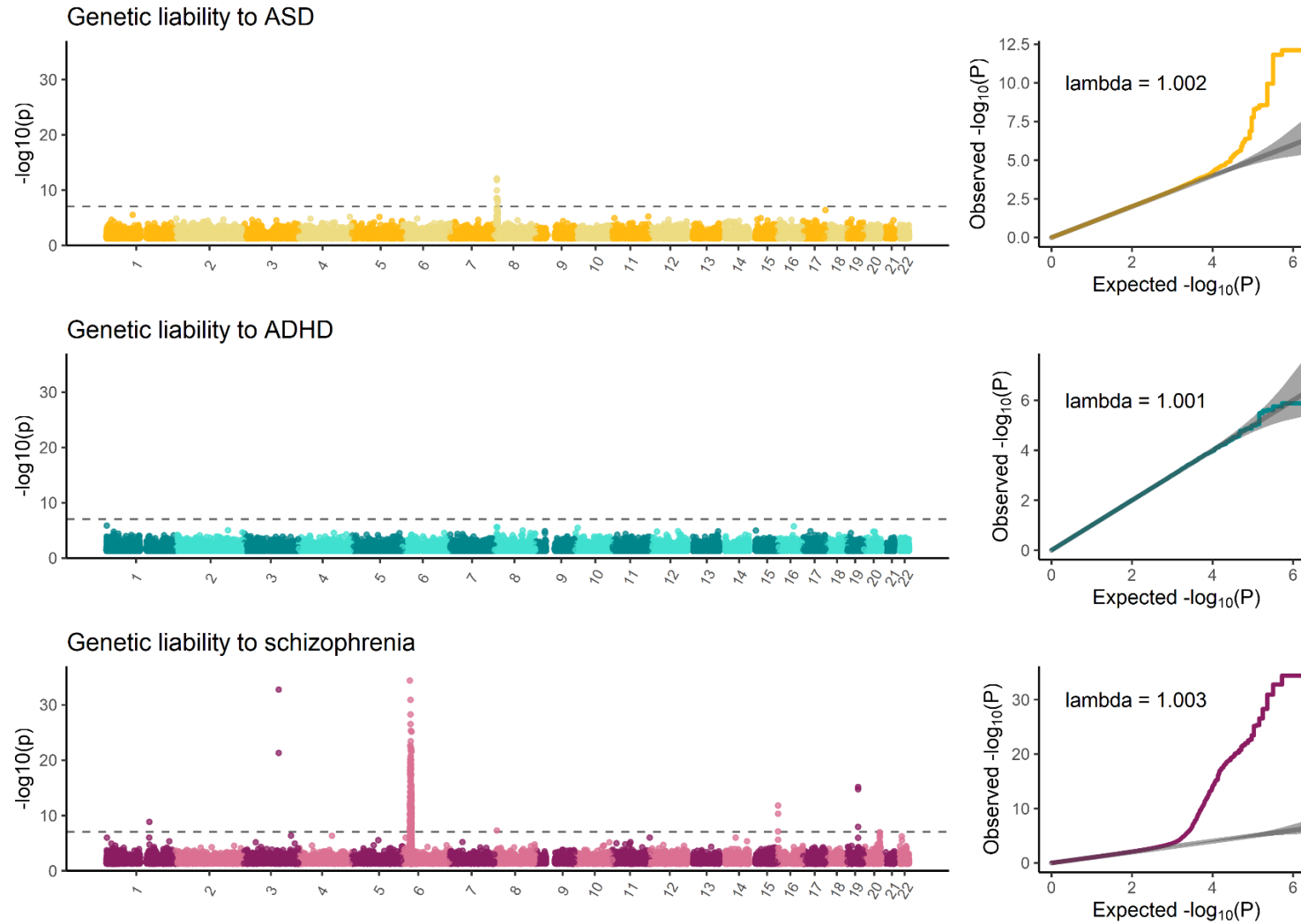
GENETIC SUSCEPTIBILITY TO NDCS AND NEONATAL DNA METHYLATION

Table 3. Top 10 significant loci of genetic susceptibility for neurodevelopment conditions with cord blood DNA methylation.

	Chr	Position	Effect	SE	<i>p</i>	Direction	Gene	Array
Genetic liability to ASD								
cg00128506	8	8641298	-0.0262	0.0037	7.7×10^{-13}	?-????-?		EPIC
cg19034770	8	8641253	-0.0255	0.0036	1.5×10^{-12}	?-????-?		EPIC
cg07513992	8	8646873	-0.0080	0.0012	1.1×10^{-10}	?-????-?	<i>MFHAS1</i>	EPIC
cg17263012	8	8646938	-0.0016	0.0003	2.7×10^{-9}	?-????-?	<i>MFHAS1</i>	EPIC
cg14469574	8	8365789	0.0088	0.0015	2.9×10^{-9}	?+????++		EPIC
cg16423305	8	8377965	0.0099	0.0017	4.2×10^{-9}	?+????++		EPIC
cg12568669	8	11666485	0.0058	0.0010	5.0×10^{-9}	+++++++	<i>FDFT1</i>	450K
cg15426459	8	11813211	-0.0041	0.0007	1.7×10^{-8}	--+-----		450K
cg05986838	8	8643530	-0.0040	0.0008	1.3×10^{-7}	?-????-?	<i>MFHAS1</i>	EPIC
cg16664915	8	10907788	-0.0039	0.0008	3.9×10^{-7}	---+----	<i>XKR6</i>	450K
Genetic liability to ADHD (none significant at epigenome-wide significance level ($p < 9 \times 10^{-8}$))								
cg25063151	1	1290554	0.0008	0.0002	1.3×10^{-6}	+++++++	<i>MXRA8</i>	450K
cg10542509	16	57689871	-0.0014	0.0003	1.8×10^{-6}	?-????+-	<i>ADGRG1</i>	EPIC
cg15741354	8	8146129	-0.0044	0.0009	2.4×10^{-6}	-----		450K
cg11548083	8	10208156	0.0033	0.0007	2.7×10^{-6}	++++++?	<i>MSRA</i>	450K
cg22805294	10	11420172	0.0005	0.0001	3.3×10^{-6}	?+????++		EPIC
cg12006072	2	186609566	0.0013	0.0003	9.1×10^{-6}	+++++++		450K
cg22727641	8	101157742	-0.0017	0.0004	1.0×10^{-6}	--+-----	<i>FBXO43</i>	450K
cg10073842	15	23894198	-0.0022	0.0005	1.0×10^{-5}	-----	<i>MAGEL2</i>	450K
cg13562353	9	34662773	0.0015	0.0004	1.3×10^{-5}	+++++?-?	<i>CCL27</i>	450K
cg27239144	12	100661541	-0.0005	0.0001	1.4×10^{-5}	-----+--	<i>SCYL2</i>	450K
Genetic liability to schizophrenia								
cg14345882	6	26364793	-0.0124	0.0010	3.9×10^{-35}	-----	<i>BTN3A2</i>	450K
cg08344181	3	125677491	0.0106	0.0009	1.7×10^{-33}	+++++++		450K
cg00903577	6	28831109	0.0171	0.0015	1.2×10^{-31}	+++?++++		450K
cg19918208	6	28828364	0.0110	0.0010	5.2×10^{-29}	+++?++++		450K
cg16677399	6	28830902	0.0105	0.0010	2.9×10^{-27}	+++?++++		450K
cg16690356	6	29945948	0.0156	0.0015	4.7×10^{-26}	+++?++++	<i>HCG9</i>	450K
cg07180897	6	32729130	0.0241	0.0023	7.0×10^{-26}	+++?++++	<i>HLA-DQB2</i>	450K
cg07134666	6	29648400	0.0331	0.0033	3.7×10^{-24}	+++?++++		450K
cg12623302	6	28058802	-0.0146	0.0015	2.2×10^{-23}	-----	<i>ZSCAN12L1</i>	450K
cg12914966	6	28830789	0.0105	0.0011	3.2×10^{-23}	+++?++++		450K

Note. Chr = chromosome; Effect = reflects proportion increase in methylation [range 0-1] for each standard deviation increase in genetic susceptibility; SE = standard error; Direction: direction of effect per cohort, in which + signifies a positive effect estimate, - signifies a negative effect estimate, and ? refers to a lack of methylation data on that site in the cohort (cohorts are listed as follows: GenR_{450K}, GenR_{EPIC}, PREDO, ALSPAC, MoBa-1, MoBa-2, MoBa-4, MoBa-8). Of note, most ASD hits are only available on the EPIC array ($n=2,504$).

Figure 1. Manhattan plots and related quantile-quantile plots. Manhattan plots show which CpGs are associated with genetic susceptibility to neurodevelopmental conditions, with the grey dotted line indicating the epigenome-wide significance of $p < 9 \times 10^{-8}$.



GENETIC SUSCEPTIBILITY TO NDCS AND NEONATAL DNA METHYLATION

Cross-NDC comparisons

ASD-PGS and ADHD-PGS ($r=.18$) were more strongly correlated with each other than with SCZ-PGS ($r=.08-.09$, **Figure 2, panel A**). Further, while probe-level methylation patterns of genetic susceptibility for NDCs were largely unrelated, three suggestive sites were shared between the ASD-PGS and ADHD-PGS (cg19034770, cg15741354, and cg11548083) but not with SCZ-PGS (**Figure 2, panel B**). A considerable proportion of variance in methylation patterns across PGSs could be attributed to heterogeneity (mean heterogeneity $>74.1\%$; **Figure 2, panel C, Supplementary Table 8**).

Phenotypic associations

In our target sample (GenREPIC, $n=1,097$), PGS and MPS together explained up to 2.7% of variance in phenotypes above covariates. On average, the MPS contributed to 37.2% of their total explained variance. ASD-PGS showed Bonferroni-corrected associations above covariates with 4/130 outcomes ($p<8\times10^{-4}$; nominally at $p<0.05$ with 20/130), ADHD-PGS with 11/130 outcomes (nominally 43/130), and SCZ-PGS with 1/130 outcomes (nominally 38/130), all above covariates. The MPS did not show any Bonferroni-corrected associations above covariates and PGS. Nominal increased variance ($p<0.05$) was found for (i) ASD-PGS-MPS in ADHD symptoms, IQ, and number sequencing abilities, (ii) ADHD-PGS-MPS in attention and total emotional and behavioural problems, ADHD diagnosis, language comprehension, number sequencing, height, and IQ, (iii) SCZ-PGS-MPS in gross motor skills (**Figure 3**). Detailed results are shown in **Supplementary Table 9-11 and Figure S3**.

Discussion

We examined whether genetic susceptibility to NDCs is associated with DNAm patterns in cord blood, pooling data from four population-based cohorts totalling almost 6,000 participants. Our meta-analytic EWAS revealed strong probe-level associations between genetic susceptibility to schizophrenia and DNAm (246 hits), compared to ASD (8 hits) and ADHD (none). SCZ-PGS hits were mainly located within the MHC on chromosome 6, a well-established genetic risk locus for schizophrenia with a key role in immune function. In contrast to probe-level hits, region-level analyses (combining proximal DNAm sites into a smaller set of DNAm regions) detected a large number of differentially methylated regions across all three PGS (130-166 regions). PGS showed little overlap in their DNAm associations, suggesting largely distinct epigenetic signals. Finally, a DNAm-based score of genetic liability nominally increased variance in several developmental outcomes.

Our findings suggest that part of the polygenic contribution for schizophrenia is detectable epigenetically in the general population, even at a modest sample size relative to the case-control GWAS used to calculate the SCZ-PGS ($n>200,000$). Observing a strong signal already at birth (preceding by many years symptom onset) provides additional support for an early-origins perspective of schizophrenia.² Previous work has found that genes implicated in schizophrenia are highly expressed in the foetal brain¹⁷; we extend this by showing that associations are also present early on a gene regulatory level (i.e., epigenetics). Perhaps unsurprisingly, our EWAS findings aligned closely with the discovery GWAS.¹⁷ The majority of identified hits clustered on chromosome 6, with 62% situated within the MHC (including our top hit cg14345882, $p=3.9\times10^{-35}$, annotated to the promoter region of *BTN3A2*), likely because of the strong and extended linkage disequilibrium pattern of this region. The MHC region is the strongest known genetic risk locus for schizophrenia,²² playing a key role in immune function (e.g., encoding proteins for antigen presentation),²³ alongside neurodevelopmental and brain-

Figure 2. Cross-NDC comparisons for top-hits. Panel A shows correlations across genetic susceptibility (PGSs) for neurodevelopmental conditions (meta-analyzed in a fixed-effect model). Panel B shows the suggestive CpGs that are shared across conditions ($p < 5 \times 10^{-5}$). Panel C shows heterogeneity (I^2) for suggestive sites. I^2 indicates the percentage of variability in associations between genetic susceptibility and DNA methylation, of a given CpGs, across two conditions, that can be attributed to variability between genetic susceptibility to the two conditions. A lower I^2 value therefore reflects more similar effect sizes, for a given CpG, across the compared conditions.

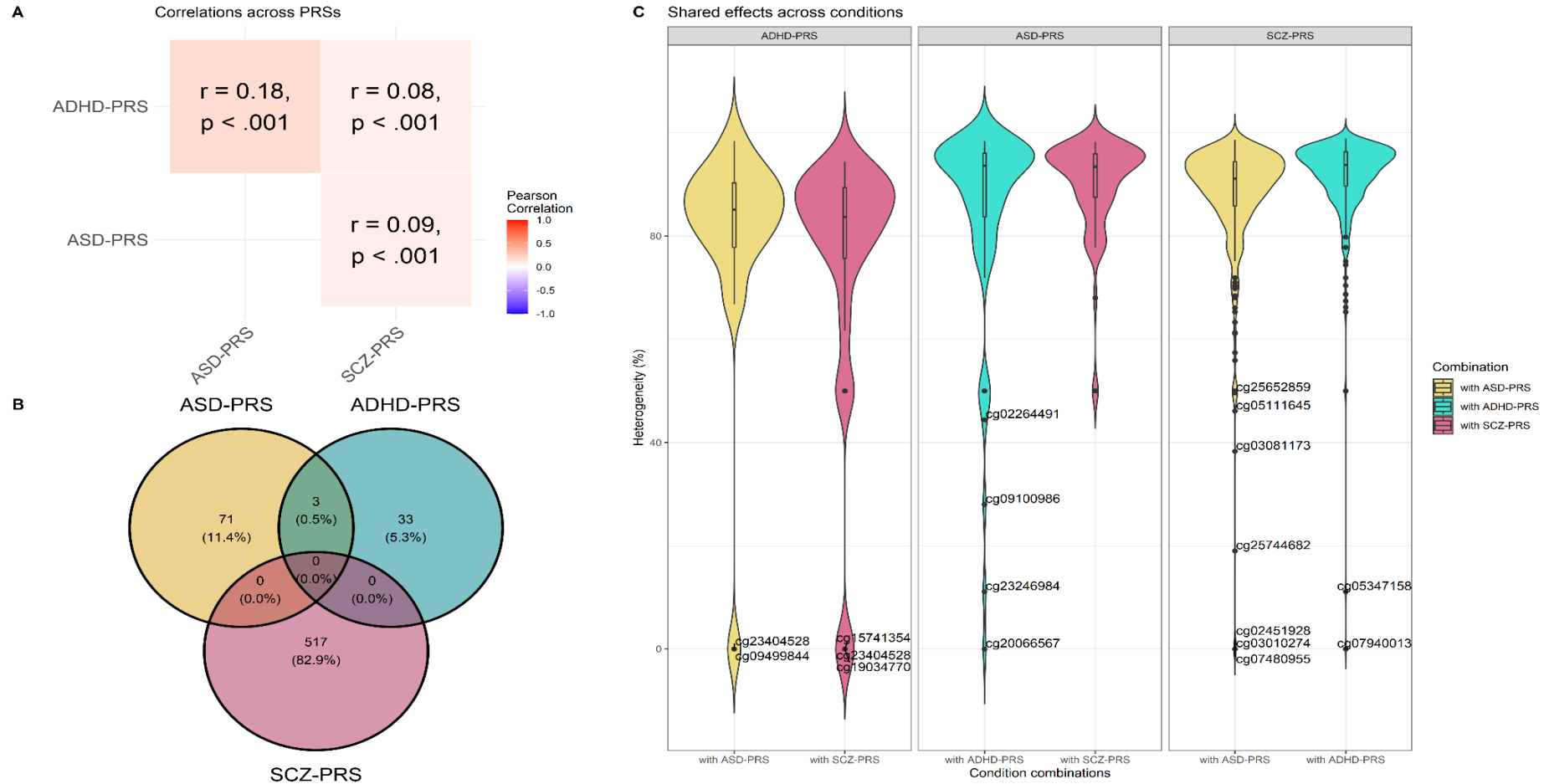
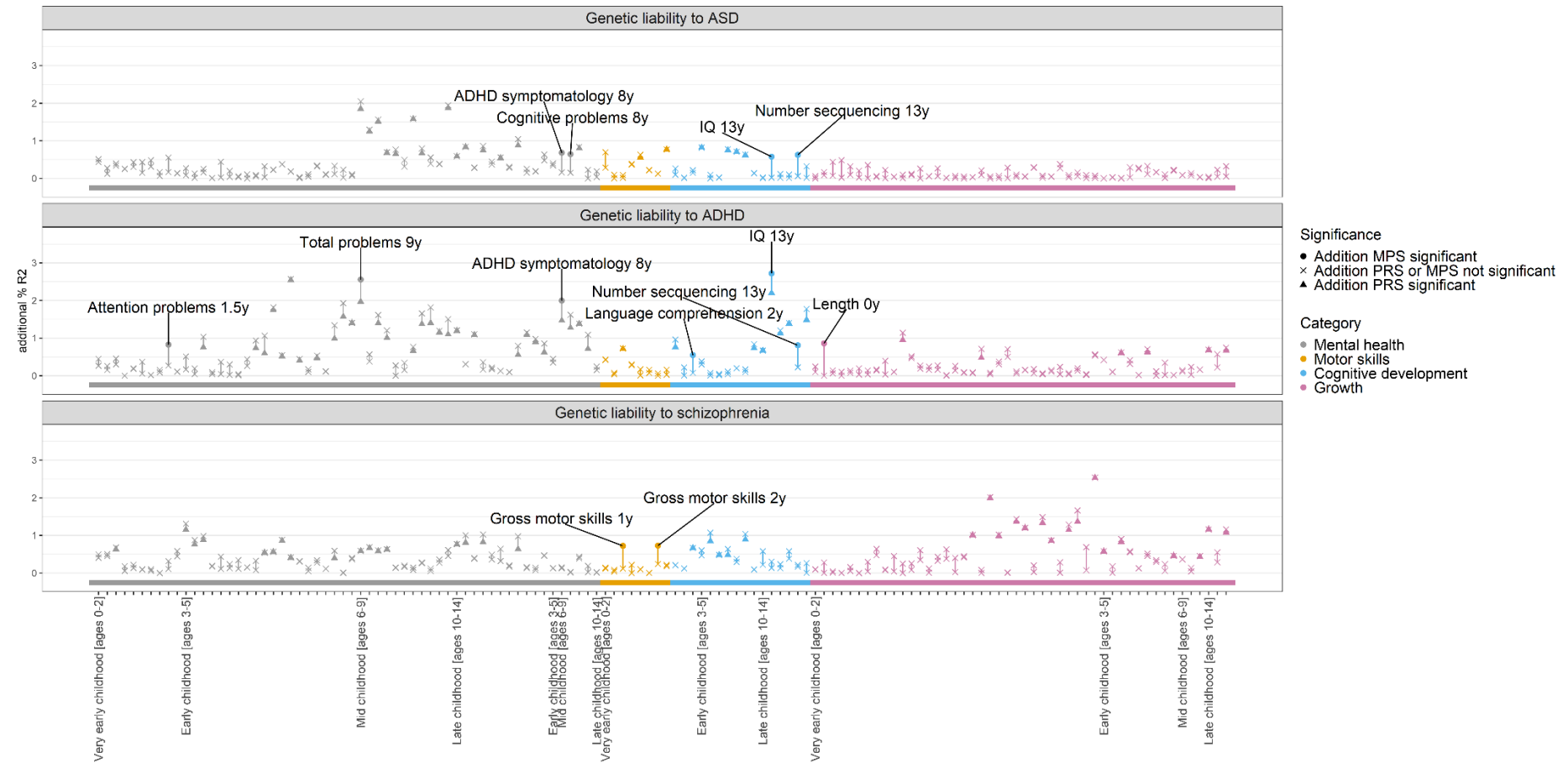


Figure 3. Phenotypic associations with genetic susceptibility (PGSs) to neurodevelopmental conditions and their methylation profile scores (MPS) in the target sample (GenREPIC; $n=1,097$). Instances where the MPS explained additional phenotypic variance on top of the PGS ($p<0.05$) are indicated. More detail can be found in Supplementary Table S9-11.



related processes (e.g., synaptic pruning).²² Other EWAS signals overlapping with the schizophrenia GWAS included significant associations of SCZ-PGS with CpGs located near (i) the *ROPN1B* in chromosome 3, an important gene for male fertility and sperm motility; (ii) the *SSTR5-AS1* gene in chromosome 16, implicated in cancer development and progression;²⁴ and (iii) the *RYR1* gene in chromosome 19, which encodes a ryanodine receptor found in skeletal muscle, of which mutations have emerged as a common cause of inherited neuromuscular disease.²⁵ Notably, while our results reveal a widespread DNAm signal, consistent with the complex genetic architecture of schizophrenia, not all GWAS peaks were mirrored in our EWAS findings. It will be of interest in future to establish whether these may be less linked to DNAm patterns generally, whether associations may emerge later in development, or whether these are present in different tissues (e.g., brain).

Substantially fewer hits were identified in the same sample for ASD-PGS and ADHD-PGS, which may reflect limited power of the PGSs due to the smaller sample size of the discovery GWAS (ASD: 18,381 individuals diagnosed and 27,969 controls; ADHD: 38,691 individuals diagnosed and 186,843 controls; schizophrenia: 76,755 individuals diagnosed and 243,649 controls),¹⁷⁻¹⁹ Interestingly, GWAS sample sizes do not entirely explain our pattern of findings, as we identified no hits for ADHD-PGS but 8 hits for ASD-PGS (lowest GWAS sample size). All ASD-PGS hits were located on chromosome 8 (1.0-1.9 Mb located from most prominent ASD GWAS loci *C8orf74*, *SOX7*, *PINX1*), on/ close to *FDFT1*, a gene implicated in the biosynthesis of cholesterol,²⁶ and *MFHAS1*, involved in modulating the innate immune system.²⁷ Unlike the stark differences in signal observed at a probe-level, we identified a similarly large number of regions (combining multiple probes) across all three PGS. This could suggest that PGS-ASD and PGS-ADHD also associate with neonatal DNAm, but that the epigenetic signal is more diffuse or that PGSs for these conditions are currently underpowered for the detection of individual CpGs. In addition, SCZ-PGS CpGs potentially were more discoverable in EWAS, as they showed higher heritability in twin-based studies, more mQTLs, and stronger associations with expression levels in blood, as compared to ASD-PGS and ADHD-PGS hits. Overall, the probe-level epigenetic signals exhibited little overlap across NDC PGSs. This may seem surprising given the known genetic (and phenotypic) correlations between ASD, ADHD, and schizophrenia. However, the magnitude of genetic correlations tends to be modest, as also observed in our data, leaving far more variance that is ‘unique’ to each condition.

We found that incorporating a DNAm-based measure of genetic susceptibility on top of PGS could considerably (yet nominally) increase explained variance in developmental outcomes. While this finding is preliminary and in need of replication, it supports continued examination into the utility of integrating information on genetic susceptibility at multiple biological layers (PGS, DNAm-based measure of genetic susceptibility) to enhance performance of early risk prediction models. It might be particularly interesting to focus on child attentional, cognitive and motor outcomes rather than emotional/behavioural symptoms or general growth parameters, as these showed largest increases in explained variance in comparison to other phenotypes. Potentially, such phenotypes are more detectable due to biological differences (i.e., a more pronounced foetal neurodevelopmental component) or measurement-related factors, for example lower measurement error and reporting bias. In addition, DNAm-based measures of genetic susceptibility to ASD and ADHD both associated prospectively with ADHD-related phenotypes, whereas the DNAm-based measures of genetic susceptibility to schizophrenia only explained additional variance in early gross motor abilities. This is consistent with previous evidence showing that PGS-SCZ correlates poorly with psychiatric symptoms in childhood,²⁸ but associates with early motor abilities.²⁹ Of note, our target dataset enabled us to examine phenotypes from birth to age 14 years, i.e., before typical schizophrenia onset. In future, a longer follow-up will be valuable for establishing how

the neonatal DNAm-based measures of genetic susceptibility to schizophrenia may associate with relevant phenotypes (e.g. psychotic experiences) as children transition into adulthood. Importantly, only 1-10% of children will develop one of these NDCs,³⁰ which raises questions about what factors interact with genetic susceptibilities to shape phenotypic expression during development, and whether DNAm may be used in this context to improve risk stratification, as it responds dynamically to both genetic and environmental influences.

Our findings need to be interpreted in light of limitations. First, potential influences of perinatal factors cannot be completely ruled out, leaving it unclear whether our findings represent direct or indirect effects of genetic susceptibility on DNAm (e.g., genetic nurture effects). Second, as we did not explicitly investigate genetic pleiotropy, observed associations may be driven by related NDCs that were not measured (e.g., intellectual disability). Third, while we consider schizophrenia as NDC, it is not included in this category within the ICD-11 and its categorization as neurodevelopmental remains a topic of debate. Fourth, our meta-analysis lacked non-European datasets limiting the extrapolation of our results across populations, and may have been influenced by selection for example the overrepresentation of healthy newborns. Finally, while we identified numerous DNAm loci associated with genetic susceptibility to NDCs, we cannot fully ascertain their functional relevance or establish their role as a potential causal pathway to NDC pathophysiology as opposed to non-causal biomarkers for genetic susceptibility.

Conclusions

By combining genetic and epigenetic data at birth, our findings lend novel insights into the early molecular correlates of genetic susceptibility to ASD, ADHD, and schizophrenia. We identify strong associations between the PGS of schizophrenia and neonatal DNAm patterns in the general population, particularly within the MHC region, lending further support for an early-origins perspective of schizophrenia and its links to adaptive immunity. Associations of neonatal DNAm with PGS of ASD and ADHD were present but more diffuse, possibly reflecting differences in the power of the discovery GWAS. In addition, epigenetic signals at birth are largely distinct between NDC PGSs. Finally, we found preliminary evidence that the addition of methylation data may enhance genetic prediction models of neurodevelopmental outcomes.

Research in context

Evidence before this study:

Neurodevelopmental conditions (NDCs) like ASD, ADHD, and schizophrenia have a strong genetic component. Epigenetic processes, such as DNA methylation, are emerging as promising molecular systems in identifying biological markers and mediators of genetic influences on these NDCs. We searched PubMed and Google Scholar for epigenome-wide association studies of PGSs of ASD, ADHD, and schizophrenia from Jan 1, 1990, to March 1, 2024. Search terms were (“methylation” OR “DNAm” OR “epigenetics”) AND (“polygenic” OR “PGS” OR “GWAS” OR “genome wide association” OR “genetic”) AND (“ASD” OR “autism” OR “ADHD” OR “attention-deficit/hyperactivity disorder” OR “schizophrenia” OR “psychosis”), restricted to articles published in English. We identified three studies that examined the link between genetic predispositions to NDCs and peripheral blood DNA methylation patterns in clinical populations (more information in **Supplement**). We did not identify any studies that were based on the general population, pooled different datasets to identify robust signals, examined cord blood at birth, compared associations between the different NDCs, or tested the potential utility of DNAm-based measures (methylation profile scores) in the prediction of (neuro)developmental outcomes.

Added value of this study:

Our epigenome-wide meta-analysis ($n=5,802$) revealed strong associations between genetic susceptibility to schizophrenia and neonatal DNAm (246 hits), compared to ASD (8 hits) and ADHD (no hits). Most schizophrenia-PGS (SCZ-PGS) hits were found in the major histocompatibility region on chromosome 6 – a well-established genetic risk locus for schizophrenia with a key role in immune function – as well as several other loci (e.g., *ROBN1B*, *SSTR5-AS*, *RYS1*, *SLC39A1*) consistent with the complex genetic architecture of schizophrenia. These findings indicate that genetic susceptibility for schizophrenia can in part be detected at an epigenetic level in the general population at birth (i.e., pre-symptom manifestation), in a comparatively modest sample size relative to the case-control GWAS used to calculate the SCZ-PGS ($n>300,000$). The identified epigenetic signals differed substantially between NDC PGSs, suggesting unique underlying molecular pathways. Finally, while preliminary and in need for further replication, we found that incorporating DNAm-based measures of genetic susceptibility nominally increased explained variance for selected phenotypes.

Implications of all the available evidence:

Our findings show that genetic susceptibility to NDCs associates with epigenetic patterns already at birth and within the general population. This adds to the growing body of literature supporting an early-origins perspective of schizophrenia. Further, our results suggest that incorporating information on genetic susceptibility at multiple biological layers (PGS, MPS) may help to enhance performance of early risk prediction models. Future studies are needed to establish the functional significance of identified DNAm loci, their role in NDC pathophysiology, and potential as diagnostic markers.

Data sharing

Summary-level and analytical code data will be made available online upon acceptance of this manuscript.

Declarations of interest

None.

Funding

The general design of the Generation R Study is made possible by financial support from Erasmus MC, Erasmus University Rotterdam, the Netherlands Organization for Health Research and Development and the Ministry of Health, Welfare and Sport. The EWAS data were funded by a grant from the Netherlands Genomics Initiative (NGI)/Netherlands Organisation for Scientific Research (NWO) Netherlands Consortium for Healthy Aging (NCHA; project nr. 050-060-810), by funds from the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, and by a grant from the National Institute of Child and Human Development (R01HD068437). This project received funding from the European Union's Horizon 2020 research and innovation programme (733206, LIFECYCLE; 848158, EarlyCause).

The PREDO study has received funding from the Academy of Finland, EraNet, EVO and VTR (special state subsidy for health science research), University of Helsinki Research Funds, the Signe and Ane Gyllenberg foundation, the Emil Aaltonen Foundation, the Finnish Medical Foundation, the Jane and Aatos Erkko Foundation, the Novo Nordisk Foundation, the Päivikki and Sakari Sohlberg Foundation, Sigrid Juselius Foundation, the Sir Jules Thorn Charitable Trust, and , and the HiLife Fellows Programme 2023-25.

Core support for ALSPAC was provided by the UK Medical Research Council and Wellcome (Grant ref: 217065/Z/19/Z) and the University of Bristol. GWAS data was generated by Sample Logistics and Genotyping Facilities at Wellcome Sanger Institute and LabCorp (Laboratory Corporation of America) using support from 23andMe. This publication is the work of the authors and they will serve as guarantors for the contents of this paper. A comprehensive list of grants funding is available on the ALSPAC website (). EW also received funding from the National Institute of Mental Health of the National Institutes of Health (award number R01MH113930) and from UK Research and Innovation (UKRI) under the UK government's Horizon Europe / ERC Frontier Research Guarantee [BrainHealth, grant number EP/Y015037/1].

The Norwegian Mother, Father and Child Cohort Study is supported by the Norwegian Ministry of Health and Care Services and the Ministry of Education and Research. We thank the Norwegian Institute of Public Health (NIPH) for generating high-quality genomic data. This research is part of the HARVEST collaboration, supported by the Research Council of Norway (#229624). We also thank the NORMENT Centre for providing genotype data, funded by the Research Council of Norway (#223273), South East Norway Versjon 6.9 3 Health Authorities and Stiftelsen Kristian Gerhard Jebsen. We further thank the Center for Diabetes Research, the University of Bergen for providing genotype data and performing quality control and imputation of the data funded by the ERC AdG project SELECTIONPREDISPOSED, Stiftelsen Kristian Gerhard Jebsen, Trond Mohn Foundation, the Research Council of Norway, the Novo Nordisk Foundation, the University of Bergen, and the Western Norway Health Authorities.

The work of C.A.M.C. is supported by the European Union's Horizon 2020 Research and Innovation Programme (EarlyCause; grant agreement No 848158), the European Union's HorizonEurope Research and Innovation Programme (FAMILY, grant agreement No 101057529; HappyMums, grant agreement No 101057390) and the European Research Council (TEMPO; grant agreement No 101039672). This research was conducted while CAMC was a Hevolution/AFAR New Investigator Awardee in Aging Biology and Geroscience Research. The work of AN is also supported by the European Union's HorizonEurope Research and Innovation Programme (FAMILY, grant agreement No 101057529) and the European Research Council (TEMPO; grant agreement No 101039672). JBP and NC are supported by the ERC under the European Union's Horizon 2020 research and innovation programme (iRISK; grant agreement No 863981).

The financial supporters did not influence the results of this article. The funders had no role in the study design, data collection, analysis, interpretation of the data, or writing of the report.

Acknowledgements

The Generation R Study is conducted by Erasmus MC, University Medical Center Rotterdam in close collaboration with the School of Law and Faculty of Social Sciences of the Erasmus University Rotterdam, the Municipal Health Service Rotterdam area, Rotterdam, the Rotterdam Homecare Foundation, Rotterdam and the Stichting Trombosedienst & Artsenlaboratorium Rijnmond (STAR-MDC), Rotterdam. We gratefully acknowledge the contribution of children and parents, general practitioners, hospitals, midwives and pharmacies in Rotterdam. The generation and management of the Illumina 450K methylation array data (EWAS data) for the Generation R Study was executed by the Human Genotyping Facility of the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, the Netherlands. We thank Mr. Michael Verbiest, Ms. Mila Jhamai, Ms. Sarah Higgins, Mr. Marijn Verkerk and Dr. Lisette Stolk for their help in creating the EWAS database. We thank Dr. A. Teumer for his work on the quality control and normalization scripts. We also thank all the research nurses, research assistants, and laboratory personnel involved in the PREDO study. We are also extremely grateful to all the families who took part in the ALSPAC study, as well as for the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses. We thank all the children and their parents for participation. We are grateful to all the participating families in Norway who take part in this on-going cohort study, MoBa. Finally, we would like to thank prof. dr. Roel Ophoff for his helpful input on the study.

References

1. World Health Organization (WHO). International Classification of Diseases, Eleventh Revision (ICD-11). <https://icdwho.int/browse11>; 2019/2021.
2. Owen MJ, O'Donovan MC. Schizophrenia and the neurodevelopmental continuum: evidence from genomics. *World Psychiatry* 2017; **16**(3): 227-35.
3. Sandin S, Lichtenstein P, Kuja-Halkola R, Hultman C, Larsson H, Reichenberg A. The heritability of autism spectrum disorder. *Jama* 2017; **318**(12): 1182-4.
4. Faraone SV, Larsson H. Genetics of attention deficit hyperactivity disorder. *Molecular psychiatry* 2019; **24**(4): 562-75.
5. Hilker R, Helenius D, Fagerlund B, et al. Heritability of schizophrenia and schizophrenia spectrum based on the nationwide Danish twin register. *Biological psychiatry* 2018; **83**(6): 492-8.
6. Wu Y, Cao H, Baranova A, et al. Multi-trait analysis for genome-wide association study of five psychiatric disorders. *Translational psychiatry* 2020; **10**(1): 1-11.
7. Aref-Eshghi E, Kerkhof J, Pedro VP, et al. Evaluation of DNA methylation epigenatures for diagnosis and phenotype correlations in 42 Mendelian neurodevelopmental disorders. *The American Journal of Human Genetics* 2020; **106**(3): 356-70.
8. Hannon E, Schendel D, Ladd-Acosta C, et al. Elevated polygenic burden for autism is associated with differential DNA methylation at birth. *Genome medicine* 2018; **10**(1): 1-13.
9. Mooney MA, Ryabinin P, Wilmot B, Bhatt P, Mill J, Nigg JT. Large epigenome-wide association study of childhood ADHD identifies peripheral DNA methylation associated with disease and polygenic risk burden. *Translational Psychiatry* 2020; **10**(1): 1-12.
10. Hannon E, Dempster E, Viana J, et al. An integrated genetic-epigenetic analysis of schizophrenia: evidence for co-localization of genetic associations and differential DNA methylation. *Genome biology* 2016; **17**(1): 1-16.
11. Cecil CA, Nigg JT. Epigenetics and ADHD: reflections on current knowledge, research priorities and translational potential. *Molecular diagnosis & therapy* 2022; **26**(6): 581-606.
12. Lewis CM, Vassos E. Polygenic risk scores: from research tools to clinical instruments. *Genome medicine* 2020; **12**(1): 1-11.
13. Girchenko P, Lahti M, Tuovinen S, et al. Cohort profile: prediction and prevention of preeclampsia and intrauterine growth restriction (PREDO) study. *International journal of epidemiology* 2017; **46**(5): 1380-1g.
14. Kooijman MN, Kruithof CJ, van Duijn CM, et al. The Generation R Study: design and cohort update 2017. *European Journal of Epidemiology* 2016; **31**(12): 1243-64.
15. Boyd A, Golding J, Macleod J, et al. Cohort profile: the 'children of the 90s'—the index offspring of the Avon Longitudinal Study of Parents and Children. *International journal of epidemiology* 2013; **42**(1): 111-27.
16. Magnus P, Birke C, Vejrup K, et al. Cohort profile update: the Norwegian mother and child cohort study (MoBa). *International journal of epidemiology* 2016; **45**(2): 382-8.
17. Trubetskoy V, Pardiñas AF, Qi T, et al. Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature* 2022; **604**(7906): 502-8.
18. Demontis D, Walters B, Athanasiadis G, et al. Genome-wide analyses of ADHD identify 27 risk loci, refine the genetic architecture and implicate several cognitive domains. *medRxiv* 2022.
19. Grove J, Ripke S, Als TD, et al. Identification of common genetic risk variants for autism spectrum disorder. *Nature genetics* 2019; **51**(3): 431-44.
20. Gervin K, Salas LA, Bakulski KM, et al. Systematic evaluation and validation of reference and library selection methods for deconvolution of cord blood DNA methylation data. *Clinical epigenetics* 2019; **11**(1): 1-15.
21. Mansell G, Gorrie-Stone TJ, Bao Y, et al. Guidance for DNA methylation studies: statistical insights from the Illumina EPIC array. *BMC genomics* 2019; **20**(1): 1-15.
22. McAllister AK. Major histocompatibility complex I in brain development and schizophrenia. *Biological psychiatry* 2014; **75**(4): 262-8.
23. Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. T cells and MHC proteins. *Molecular Biology of the Cell* 4th edition: Garland Science; 2002.
24. Wang B, Zhao L, Chi W, Cao H, Cui W, Meng W. Aberrant methylation-mediated downregulation of lncRNA SSTR5-AS1 promotes progression and metastasis of laryngeal squamous cell carcinoma. *Epigenetics & Chromatin* 2019; **12**(1): 1-17.
25. Snoeck M, Van Engelen B, Küsters B, et al. RYR 1-related myopathies: a wide spectrum of phenotypes throughout life. *European journal of neurology* 2015; **22**(7): 1094-112.
26. Goldstein JL, Brown MS. Regulation of the mevalonate pathway. *Nature* 1990; **343**(6257): 425-30.
27. Dihanich S. MASL1: a neglected ROCO protein. *Biochemical Society Transactions* 2012; **40**(5): 1090-4.

28. Hughes DE, Kunitoki K, Elyounssi S, et al. Genetic patterning for child psychopathology is distinct from that for adults and implicates fetal cerebellar development. *Nature neuroscience* 2023; **26**(6): 959-69.
29. Serdarevic F, Jansen PR, Ghassabian A, et al. Association of genetic risk for schizophrenia and bipolar disorder with infant neuromotor development. *JAMA psychiatry* 2018; **75**(1): 96-8.
30. National Institute of Mental Health. Statistics. April 2023 2023.
<https://www.nimh.nih.gov/health/statistics> (accessed June 14 2023).